

Abstract Title Page
Not included in page count.

Title:

Do Schools Respond To Pressure? Evidence from NCLB Implementation Details

Authors and Affiliations:

Vivian C Wong, University of Virginia

Coady Wing, Indiana University

David Martin, University of Virginia

Abstract Body

Background / Context:

Over the last decade, accountability reform has been at the forefront of the domestic policy agenda. Although the Obama Administration was critical of some elements of No Child Left Behind (NCLB), its policies endorsed high-stakes testing and expanded the scope of the stakes. With the Race to the Top and an NCLB waiver process, the administration doubled down on using student test results for high stakes purposes, making not only schools but also teachers and principals accountable for student achievement growth. Recently, the House and Senate passed bills to replace NCLB. In both versions of the bill, the centerpiece of NCLB remained intact – identifying failing schools through standardized tests of students. What remains in contention, however, are policy levers that states have to identify and address failing schools. On the one side, there is a push to provide states with greater flexibility in establishing their own learning standards and accountability sanctions and, on the other, a strong federal mandate to ensure that states hold schools accountable for improving student achievement and reducing inequality.

Responses to the original NCLB provision can provide useful empirical evidence for informing the current policy debate over accountability reform. Under NCLB, states were required to have 100% proficiency among students by 2014, but states also had significant flexibility in the implementation of their accountability policies. Over the last decade, few empirical studies have documented systematically states' implementation of accountability policies, and the impact of these policies on student outcomes. A key methodological challenge has been the lack of quantitative measures for summarizing the multiple policy levers that states had under NCLB – from selecting test assessments, to establishing proficiency cutoffs, to providing exemption rules that effectively lowered standards for most schools in the state. Moreover, although states had discretion over the details of implementation under NCLB, all states were subjected to the same federal requirements, making it challenging for policy researchers to find appropriate comparisons for assessing policy effects.

Purpose / Objective / Research Question / Focus of Study:

This study addresses methodological challenges for evaluating NCLB by introducing a new quantitative measure for describing states' accountability systems. To create the implementation measure of states' accountability policies from 2003 to 2011, we combine a dataset we created of states' accountability policies with information from several federal data sources, including the NAEP and the Common Core of Data (CCD). Our implementation measure is unique in that it depends only on state policies, but not on population characteristics of schools and students within states. The measure allows us to describe quantitatively states' implementation of accountability policies during the NCLB pre-waiver period, to assess how these policies changed over time, and to examine how schools responded to state accountability pressures.

Setting:

This is a national evaluation of states' implementation of No Child Left Behind from 2003 to 2011.

Population / Participants / Subjects:

In our study, we examine states' stringency in accountability policies from 2003 to 2011, as well as how schools responded to states' accountability policies from the same time period. As such, the populations of interest are represented in our study, which include all fifty states in the

United States, and the percentage of schools that made Adequate Yearly Progress in each state from 2003 to 2011.

Intervention / Program / Practice:

In 2002, President George W. Bush signed the federal accountability policy No Child Left Behind into law. The goal is to align teaching and learning practices such that all students are “proficient” by 2014. Under NCLB’s pre-waiver period, states set annual targets (in percentage of students proficient) to help schools and districts meet the 2014 federal mandate. However, states also had discretion over at least three important implementation decisions. First, they were allowed to select the measures used to assess students’ proficiency. Second, they could determine the steepness of the improvement trajectories that schools were required to follow. Third, they were allowed to introduce so-called “exemption rules” that include confidence interval and safe harbor rules. These policies effectively lowered the performance requirements for many schools. As such, although state proficiency requirements became more stringent over time, “exemption rules” also provided schools and districts with outlets to reduce accountability pressures. Combined, the ratcheting up of proficiency requirements as well as the inclusion of exemption rules introduced tremendous variation in the stringency of AYP policies across states and time.

Researchers have made efforts to link states’ NCLB implementation with school and student outcomes, but the results have been mixed. For example, Davidson, Reback, Rockoff, and Schwartz (2013) examined how states’ implementation of NCLB policies affected schools’ AYP failure rates from 2003 to 2005. Overall, they found that schools’ failure rates were correlated with states’ implementation of the confidence interval and minimum subgroup size rules. Wei (2012) looked at the connection between states’ implementation of specific exemption rules and student outcomes, as measured by NAEP reading and math scores. The author developed a predictive model that linked state population characteristics with the adoption of more stringent AYP rules, and studied the correlation between these predictions and students’ NAEP scores. Overall, state NCLB stringency was associated with negative cognitive outcomes for both whites and Hispanics, but not for black students. One concern with Wei’s approach, however, is that state implementation decisions may be confounded with state population characteristics.

Research Design:

Our study employs a research design called simulated instrumental variables, which has not been used to study educational reforms but is well suited to uncover links between policy implementation and outcomes. To create the simulated stringency rates for each state during the pre-waiver period, we began by creating a database of AYP rules for each state and year from 2003 to 2011. Using AYP rule data, we then developed an “AYP calculator,” which takes the percentage of proficient students, cell sizes and other performance metrics of subgroups in schools, and returns a variable indicating whether a given school would make AYP according to each state’s rules for each year. With the calculator in hand, we constructed a measure of NCLB implementation that depended on adopted state rules but not on population characteristics of the state. We accomplished this by constructing a fixed basket of schools, and then by “feeding” these schools through the calculator to determine the percentage of schools in the fixed basket that would make AYP for the state and year. The result was a state-by-year level dataset showing the “simulated AYP” pass rates, or our measure of implementation stringency for each state and

year. Importantly, because the basket of schools did not change across states and time periods, the variation in simulated pass rates arose purely from differences in rules used to determine AYP and not on changes in the population of schools. By itself, the stringency measure provides useful descriptive information about how states responded to the NCLB federal mandate from 2003 to 2011.

Analysis:

To examine schools' responses to accountability stringency, we compare states' implementation stringency to the percentage of schools that actually failed to make AYP for each state from 2003 to 2011. Note that to aid interpretation of results, we redefine implementation stringency and school responses to *school failure rates for making AYP*.

To construct our test, we begin by conceptualizing three possible toy responses from schools under accountability pressure (Figure 1). Figure 1 shows that in a static model, as AYP becomes incrementally more stringent, the corresponding percentage of schools should fail to make AYP. This means that as stringency pressure increases by 1%, then the percentage of schools that actually fail to make AYP should also increase by 1%. On the other hand, we may observe that as accountability pressures become more stringent, fewer schools than we would expect from the static model fail AYP. This suggests that schools succeed in responding to accountability pressures. Finally, if schools are demoralized by accountability pressure -- or adopt maladaptive practices -- they may fail AYP at higher rates than what is expected under the static model.

Using our observed data, we test school responses to accountability pressures through a differences-in-differences approach with state and time fixed effects, and a stringency measure that is exogenous of school and student population characteristics. As such, we run a regression of states' actual AYP failure rates against states' simulated AYP failure rates including state and year fixed effects, such that:

$$\text{Ln(ActualAYPFail)}_{st} = \beta_0 + \beta_1 \text{Ln(SimulatedAYPFail)}_{st} + \theta_s + \delta_t + \varepsilon_{st}$$

Here, we use the natural log of states' actual AYP school fail rates and states' simulated AYP fail rates to approximate the relationship between the independent and dependent variables. The Null hypothesis of interest is $\beta_1=1$, which states that if stringency in AYP policy and actual AYP fail rates perfectly correspond to each other, schools do not respond to accountability pressures. However, a slope less than one suggest that schools maintain their performance despite tougher AYP requirements, and a slope greater than one indicates that schools fail at a higher rate than expected.

Findings / Results:

In our preliminary analyses, we used the population of Pennsylvania schools in 2007-2008 to serve as our fixed sample that was "fed" through the AYP calculator. We chose Pennsylvania schools because the state department of education provided us with sufficient input information needed for our calculator and included schools with enough variation that reflect changes in state policies across time.

Figure 2 provides comparisons of simulated AYP pass rates with actual AYP pass rates by state (Table 1 presents a comparison of results). Here, the X-axis depicts years during the pre-waiver period in NCLB, and the Y-axis depicts percentage of schools that actually make AYP.

The green line describes states' expected school AYP pass rates based purely on state policies and not on population characteristics of schools. In other words, the green line summarizes our measure of stringency for each state from 2003 to 2011. Positive slopes from 2003 to 2005 show that in many states, stringency was lessened shortly after AYP policies were implemented in 2003. This was due to the introduction of exemption rules such as confidence intervals and safe harbor policies. However, negative slopes from 2005 to 2011 indicate that stringency of AYP policies increased in many states over the latter half of the period. This was due to increases in percent proficiency thresholds in order for schools to meet the 100% proficiency target in 2014.

The red line in Figure 2 depicts actual AYP pass rates in states from 2003 to 2011. Overall, the red lines show that from 2003 to 2011, fewer and fewer schools made AYP under NCLB, especially in the latter part of the period. A simple interpretation of schools' actual pass rates suggests that many schools were failing to meet the intended target of NCLB (that all students were proficient by 2014). However, a comparison of trends from our simulated pass percentage with actual AYP pass percentage suggest that in some states at least, schools did appear to respond to increasing accountability pressure by at least maintain their proficiency performance.

Figure 3 shows that in Maryland, schools responded positively to accountability pressure until about 2008. Here, even as AYP rules became increasingly stringent, the percentage of schools making AYP actually increased. After 2008, however, the slope of schools' actual AYP rates became sharply negative – much more so than even the slope of the stringency measure. This suggests that post 2008, Maryland schools became demoralized under NCLB pressure, or that they adopted changes that resulted in worse performance in AYP. For Wisconsin, AYP pressure steadily increased every two to three years from 2003-2011. Although there is a slight decline in the percentage of schools that actually make AYP over this time period, the decline is not nearly as steep as one would expect if schools did not respond to accountability pressure at all. This suggests that in Wisconsin, schools responded to more stringent AYP pressures by improving their performance on AYP. In contrast, Figure 4 shows that in Mississippi, when AYP pressures became more stringent, a corresponding percentage of schools failed to make AYP. The figure suggests that in Mississippi, schools were unable to – or did not want to – respond to increasing accountability pressure.

Finally, to understand schools' overall response to AYP pressures, we regressed states' actual AYP failure rates on the simulated AYP failure rates (with state and year fixed effects included in the model). Overall, we find that a 1% increase in stringency of AYP policy led to a .07% increase in schools' failure rates (Table 2). With the Null hypothesis that $\beta_1=1$, the result was statistically significant. This indicated that overall, schools responded positively to increased accountability pressure by attempting to meet AYP requirements.

Conclusions:

One limitation of the results that we present here is that they do not address differences in test difficulty across states. The final paper will include results from using the NAEP fixed sample, where we address differences in test difficulty across states and NAEP. We do this by programming an alternate AYP calculator that starts with input data from a NAEP fixed sample of students, and comparing their NAEP scores to NAEP equivalent scores with state proficiency standards. Our analyses indicate that our results were not sensitive to sample selection, or inclusion of the test difficulty in the state stringency measure.

Appendices

Not included in page count.

Appendix A. References

- Anderson, Butcher, & Schanzenbach. (2014). Adequate (or Adipose?) Yearly Progress: Assessing the Effect of No Child Left Behind on Children's Obesity. NBER Working Paper No. 16873.
- Carey, K. (2007). *The Pangloss index: How states game the No Child Left Behind Act*. Washington, DC: Education Sector.
- Center on Education Policy. *Lessons from the classroom level: Federal and state accountability in Rhode Island* (2008). Washington, DC: Center on Education Policy.
- Davidson, E., Reback, R., Rockoff, J.E., Schwartz, H.L. (2013). Fifty ways to leave a child behind: Idiosyncrasies and discrepancies in states' implementations of NCLB. NBER Working Paper 18988.
- Dee, T., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.
- De Mello, V.B. (2011). Mapping State Proficiency Standards Onto NAEP Scales: Variation and Change in State Standards for Reading and Mathematics, 2005-2009. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.
- De Mello, V.B. (2015). Mapping State Proficiency Standards Onto NAEP Scales: Results From the 2013 NAEP Reading and Mathematics Assessments. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.
- Figlio, D. N., & Winicki, J. (in press). Food for thought? The effects of school accountability plans on school nutrition. *Journal of Public Economics*.
- Gruber, J. & Simon, K. (2008). Crowd-out 10 years later: Have recent public insurance expansions crowded out private insurance? *Journal of Health Economics*, 27, 201-207.
- Holbein, J. & Ladd, H. F. (2015). Accountability Pressures and Non-achievement Student Behaviors. Working Paper at <http://www.caldercenter.org/sites/default/files/WP%20122.pdf>
- McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., Rojas, D., William, P., and Wolman, M. (2008). Comparison Between NAEP and State Mathematics Assessment Results: 2003 (NCES 2008-475). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

- McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., Rojas, D., William, P., and Wolman, M. (2008). Comparison Between NAEP and State Reading Assessment Results: 2003 (NCES 2008-474). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Rogasa, D. (2003). The "99% confidence" scam: Utah-style calculations. Stanford University.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2007). *Feeling the Florida heat: How low-performing schools respond to voucher and accountability pressure*. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.
- Taylor, Stecher, O'Day, et al. (2010). State and Local Implementation of the No Child Left Behind Act. Volume IX – Accountability Under NCLB: Final Report. Washington, DC: U.S. Department of Education.
- Wei, X. (2012). Are more stringent NCLB state accountability systems associated with better student outcomes? An analysis of NAEP results across states. *Educational Policy*. 26(2), 268-308.
- Wong, M., Cook, T. D., & Steiner, P. M. (2014). No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series. *Journal of Research on Educational Effectiveness*. 8(2).

Appendix B. Tables and Figures

Table 1. Simulated Versus Actual Percentages of Schools that Made AYP in 2003 and 2011

| State | 2003 | | 2011 | |
|-------|----------------|-------------|----------------|-------------|
| | Simulated Pass | Actual Pass | Simulated Pass | Actual Pass |
| AK | 54% | 42% | 40% | 46% |
| AL | 69% | 96% | 39% | 73% |
| AR | 83% | 64% | 53% | 65% |
| AZ | ** | 73% | 67% | 58% |
| CA | ** | 54% | ** | 34% |
| CO | 40% | 62% | 16% | 42% |
| CT | 58% | 85% | 36% | 53% |
| DC | ** | ** | ** | 13% |
| DE | 50% | 41% | 35% | 78% |
| FL | 50% | 15% | 24% | 9% |
| GA | 73% | 64% | 38% | 73% |
| HI | 91% | 39% | 53% | 41% |
| IA | 69% | 99% | 39% | 74% |
| ID | ** | 65% | ** | 62% |
| IL | 87% | 66% | 49% | 33% |
| IN | 67% | 77% | 49% | 51% |
| KS | 36% | 71% | 43% | 84% |
| KY | 26% | 59% | 32% | 43% |
| LA | 94% | 94% | 69% | 78% |
| MA | 64% | 54% | 26% | 18% |
| MD | 61% | 65% | 12% | 55% |
| ME | 86% | 73% | 42% | 37% |
| MI | 82% | 68% | 39% | 85% |
| MN | 74% | 82% | 23% | 45% |
| MO | 48% | 44% | 54% | 25% |
| MS | 90% | 77% | 58% | 52% |
| MT | 80% | 79% | 46% | 72% |
| NC | 44% | 47% | 39% | 28% |
| ND | 67% | 75% | 17% | 47% |
| NE | 45% | 56% | 49% | 73% |
| NH | 58% | 67% | 14% | 29% |
| NJ | 57% | 57% | 43% | 47% |
| NM | 56% | 75% | 44% | 14% |
| NV | 84% | 57% | 55% | 47% |
| NY | ** | 74% | ** | 53% |
| OH | 71% | 76% | 26% | 60% |
| OK | ** | 77% | ** | 70% |
| OR | 93% | 67% | 62% | 54% |
| PA | 81% | 65% | 63% | 75% |

| 2003 | | | 2011 | |
|-------|-------------------|----------------|-------------------|----------------|
| State | Simulated Pass | Actual Pass | Simulated Pass | Actual Pass |
| RI | 77% | 69% | 38% | 81% |
| SC | 93% | 20% | 42% | 24% |
| SD | 74% | 62% | 55% | 83% |
| TN | 47% | 52% | 73% | 51% |
| TX | 81% | 81% | 37% | 72% |
| UT | 79% | 64% | 47% | 76% |
| VA | 85% | 58% | 41% | 39% |
| VT | ** | 87% | ** | 28% |
| WA | 82% | 72% | 31% | 38% |
| WI | 83% | 96% | 59% | 89% |
| WV | ** | 57% | ** | 52% |
| WY | 88% | 80% | 67% | 93% |

Table 2. Schools' AYP responses to increasing stringency in state policies

| | Ln(Actual AYP Fail Rates) |
|--------------------------|---------------------------|
| Simulated AYP Fail Rates | .07* |
| State FE | Yes |
| Year FE | Yes |
| *p< .001 | |

Figure 1: School Responses to Increasing Stringency in AYP Policies

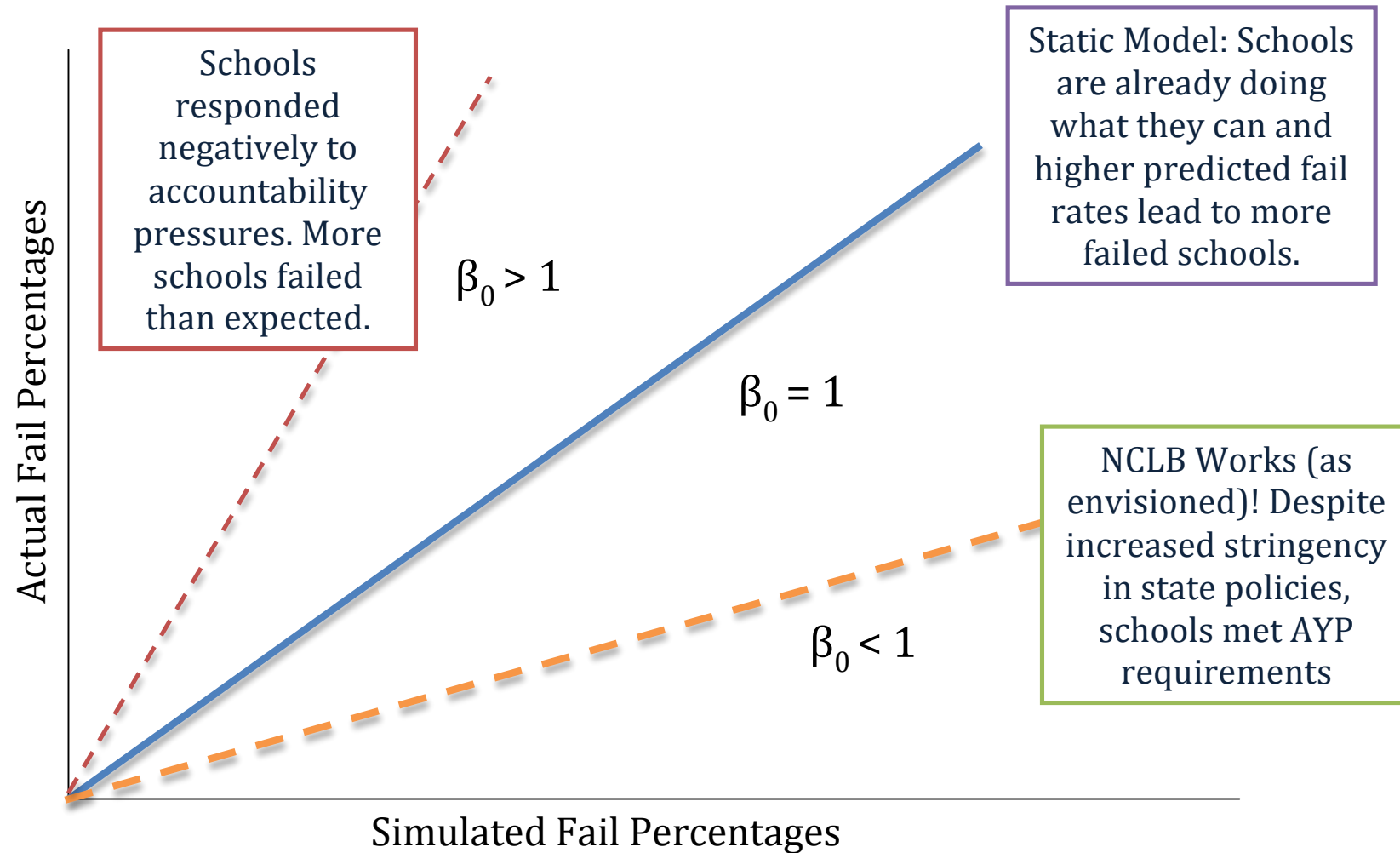
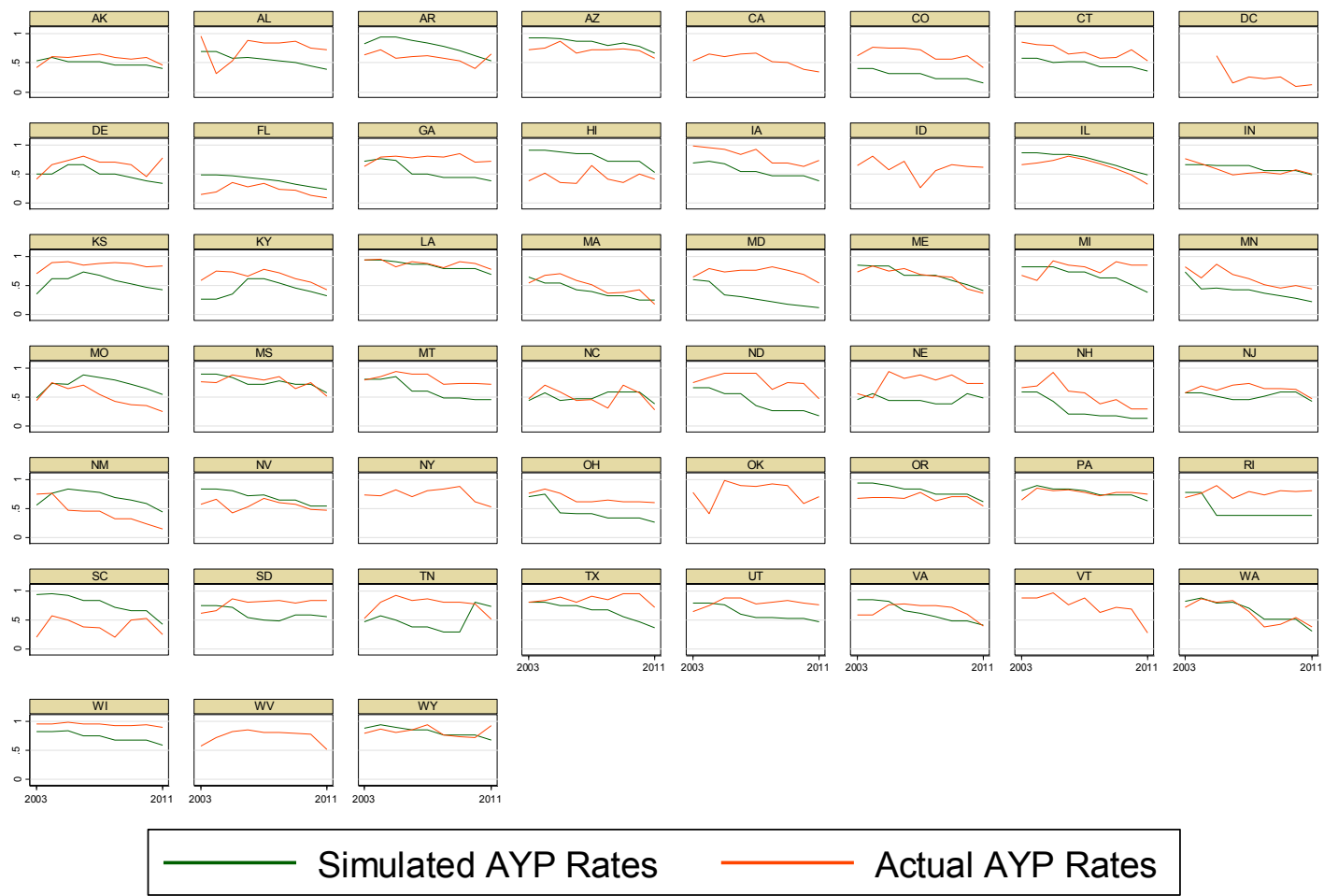


Figure 2: Summary of Simulated versus Actual AYP Pass Percentages by State



Graphs by state

Figure 3: Schools Responding to Accountability Pressures

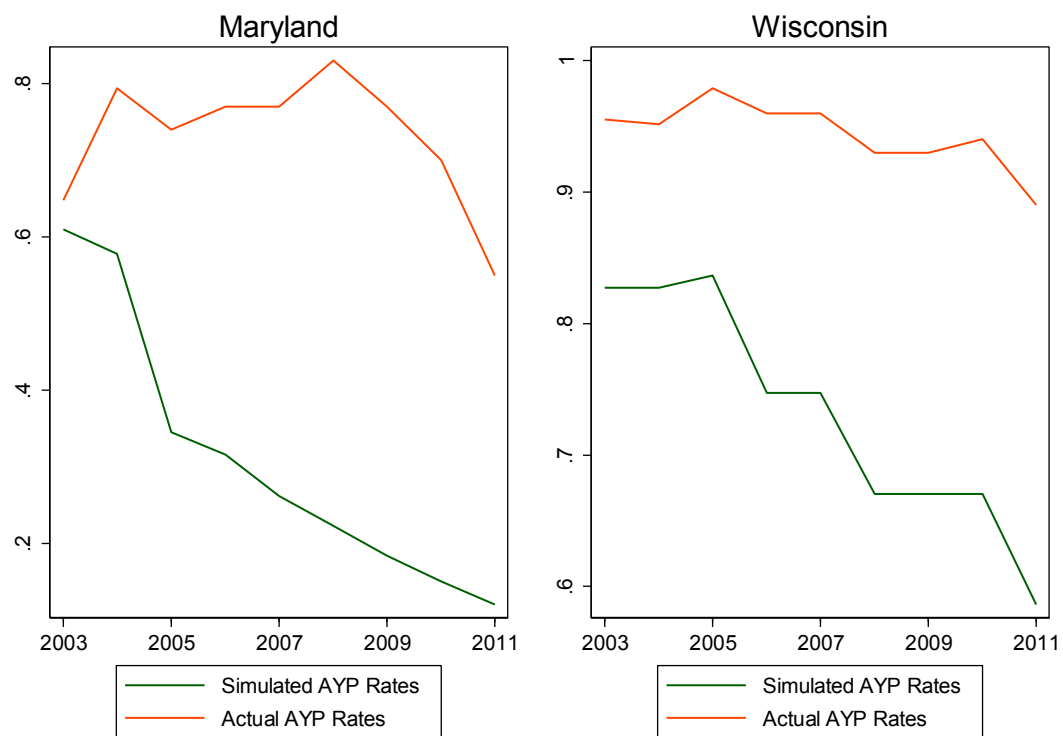


Figure 4: Schools response to AYP pressure in Mississippi

